

-1-

PROCESS FOR OXIDISING AROMATIC COMPOUNDS

The invention relates to a process for enzymatically oxidising halogenated aromatic compounds.

Chlorinated aromatic compounds such as the chlorobenzene and polychlorinated biphenyls (PCBs) are among the most wide-spread organic contaminants in the environment due to their common application as solvents, biocides, and in the heavy electrical industry. They are also some of the most problematic environmental pollutant, not only because of the health hazards (lipid solubility and hence accumulation in fatty tissues, toxicity and carcinogenicity) but also because of their slow degradation in the environment.

Whilst microorganisms have shown extraordinary abilities to adapt and evolve to degrade most of the organic chemicals released into the environment, the most chemically inert compounds such as PCBs do persist for two main reasons. First, these compounds have very low solubility in water and therefore their bioavailability is low. Research into this problem has focussed on the use of detergents and other surfactants to enhance their solubility and bioavailability. Second, these compounds require activation by enzymatic oxidation or reduction, and it can take a long time for the necessary genetic adaptations by microorganisms to occur, and even then the organisms may not be stable and viable.

We have now found, according to the present invention, that a monooxygenase, in particular P450_{cam} and its physiological electron transfer partners putidaretoxin and putidaretoxin reductase, can be used to oxidise halogenated aromatic compounds. Also mutants of the monooxygenase with substitutions in the active site have enhanced oxidation activity. Thus suitable monooxygenases can be expressed in microorganisms, animals and plants which are going to be used to oxidise the halogenated aromatic compounds.

Accordingly the present invention provides a process for oxidising a substrate which is a halo aromatic compound, which process comprises oxidising said substrate with a monooxygenase enzyme.

The process may be carried out in a cell that expresses:

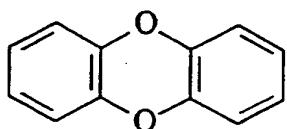
- (a) the enzyme
- (b) an electron transfer reductase; and
- (c) an electron transfer redoxin.

-2-

The halo aromatic compound is typically a benzene or biphenyl compound. The benzene ring is optionally fused and can be substituted. The halogen is typically chlorine. In many cases there is more than one halogen atom in the molecule, typically 2 to 5 or 6, for example 3. Generally 2 of the halogen atoms will be ortho or para to one another. The compound may or may not contain an oxygen atom such as a hydroxy group, an aryloxy group or a carboxy group. The compound may or may not be chlorophenol or a chlorophenoxyacetic compound.

Specific compounds which can be oxidised by the process of the present invention include 1, 2; 1,3- and 1,4-dichlorobenzene, 1, 2, 4; 1, 2, 3- and 1, 3, 5-trichlorobenzene, 1, 2, 4, 5- and 1, 2, 3, 5-tetrachlorobenzene, pentachlorobenzene, hexachlorobenzene, 3,3'-dichlorobiphenyl and 2, 3, 4, 5, 6- and 2, 2', 4, 5, 5'-pentachlorobiphenyl.

Other compounds which can be oxidised by the process include recalcitrant halo aromatic compounds, especially dioxins and halogenated dibenzofurans, and the corresponding compounds where one or both oxygen atoms is/are replaced by sulphur, in particular compounds of the formula:



which possess at least one halo substituent, such as dioxin itself, 2,3,7,8-tetrachlorodibenzodioxin.

The oxidation typically gives rise to 1, 2 or more oxidation products. These oxidation products will generally comprise 1 or more hydroxyl groups. Generally, therefore, the oxidation products are phenols which can readily be degraded. It is particularly noteworthy that pentachlorobenzene and hexachlorobenzene can be oxidised in this way since they are very difficult to degrade. In contrast the corresponding phenols can be readily degraded by a variety of Pseudomonas and other bacteria. The atom which is oxidized is generally a ring carbon.

The enzyme is typically a natural monooxygenase or a mutant thereof. The

-3-

natural monooxygenase is generally a prokaryotic or eukaryotic enzyme. Typically it is a haem-containing enzyme and/or a P450 enzyme. The monooxygenase may or may not be a TfdA (2,4-dichlorophenoxy) acetate/ α -KG dioxygenase. The monooxygenase is generally of microorganism (e.g. bacterial), fungal, yeast, plant or animal origin, typically of a bacterium of the genus *Pseudomonas*. These organisms are typically soil, fresh water or salt water dwelling. In the case of a mutant monooxygenase the non-mutant form may or may not be able to oxidize the substrate.

The monooxygenase typically has a coupling efficiency of at least 1%, such as at least 2%, 4%, 6% or more. The monooxygenase typically has a product formation rate of 10 at least 5 min⁻¹, such as at least 8, 10, 15, 20, 25, 50, 100, 150 min⁻¹ or more. The coupling efficiency or product formation rate is typically measured using any of the substrates or conditions mentioned herein. Thus they are typically measured in the in vitro conditions described in Example 2, in which case the relevant monooxygenase, reductase and redoxin would be present instead of, but at the same concentration as, P450_{cam}, putidaretoxin 15 reductase and putidaretoxin.

The mutant typically has at least one mutation in the active site. A preferred mutant comprises a substitution of an amino acid in the active site by an amino acid with a less polar side chain. Thus the amino acid is typically substituted with an amino acid which is above it in Table 1.

-4-

Table 1. HYDROPATHY SCALE FOR AMINO ACID SIDE CHAINS

5

10

15

20

25

30

35

	Side Chain	Hydropathy
	Ile	4.5
	Val	4.2
	Leu	3.8
	Phe	2.8
	Cys	2.5
	Met	1.9
	Ala	1.8
	Gly	-0.4
	Thr	-0.7
	Ser	-0.8
	Trp	-0.9
	Tyr	-1.3
	Pro	-1.6
	His	-3.2
	Glu	-3.5
	Gln	-3.5
	Asp	-3.5
	Asn	-3.5
	Lys	-3.9
	Arg	-4.5

An amino acid 'in the active site' is one which lines or defines the site in which the substrate is bound during catalysis or one which lines or defines a site through which the substrate must pass before reaching the catalytic site. Therefore such an amino acid typically interacts with the substrate during entry to the catalytic site or during catalysis. Such an interaction typically occurs through an electrostatic interaction (between charged or polar groups), hydrophobic interaction, hydrogen bonding or van der Waals forces.

The amino acids in the active site can be identified by routine methods to those skilled in the art. These methods include labelling studies in which the enzyme is allowed to bind a substrate which modifies ('labels') amino acids which contact the substrate. Alternatively the crystal structure of the enzyme with bound substrate can be obtained in order to deduce the amino acids in the active site.

-5-

The monooxygenase typically has 1, 2, 3, 4 or more other mutations, such as substitutions, insertions or deletions. The other mutations may be in the active site or outside the active site. Typically the mutations are in the 'second sphere' residues which affect or contact the position or orientation of one or more of the amino acids in the active site. The insertion is typically at the N and/or C terminal and thus the enzyme may be part of a fusion protein. The deletion typically comprises the deletion of amino acids which are not involved in catalysis, such as those outside the active site. The monooxygenase may thus comprise only those amino acids which are required for oxidation activity.

The other mutations in the active site typically alter the position and/or conformation of the substrate when it is bound in the active site. The mutation may make the site on the substrate which is to be oxidized more accessible to the haem group. Thus the mutation may be a substitution to an amino acid which has a smaller or larger, or more or less polar, side chain.

The other mutations typically increase the stability of the protein, or make it easier to purify the protein. They typically prevent the dimerisation of the protein, typically by removing cysteine residues from the protein (e.g. by substitution of cysteine at position 334 of P450_{cam}, or at an equivalent position in a homologue, preferably to alanine). They typically allow the protein to be prepared in soluble form, for example by the introduction of deletions or a poly-histidine tag, or by mutation of the N-terminal membrane anchoring sequence. The mutations typically inhibit protein oligomerisation, such as oligomerisation arising from contacts between hydrophobic patches on protein surfaces.

Typically the mutant monooxygenase is at least 70% homologous to a natural monooxygenase on the basis of amino acid identity.

Any of the homologous proteins mentioned herein are typically at least 70% homologous to a protein or at least 80 or 90% and more preferably at least 95%, 97% or 99% homologous thereto over at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous amino acids. The contiguous amino acids may include the active site. This homology may alternatively be measured not over contiguous amino acids or nucleotides but over only the amino acids in the active site.

The monooxygenase is preferably:

- (i) P450_{cam},
- (ii) a naturally occurring homologue of (i),

-6-

(iii) a mutant of (i) or (ii).

S4b B¹ Typically (i) is any allelic variant of P450_{cam} of *Pseudomonas putida* (e.g. of the polypeptide sequence shown in SEQ ID No. 1). Typically (ii) is a species homologue of (i) which has sequence homology with (i), and is typically P450_{BM-3} of *Bacillus megaterium* (e.g. the polypeptide sequence shown in SEQ ID No. 2), P450_{terp} of *Pseudomonas sp.*, P450_{eryF} of *Saccharopolyspora erythraea*, or P450 105 D1 (CYP105) of *Streptomyces griseus* strains.

The active site of (ii) or (iii) may be substantially the same as the active site of (i) or any of the mutants of (i) mentioned herein. Thus the site may comprise the same amino acids in substantially the same positions.

Typically in (iii) amino acid 96 of P450_{cam}, or the equivalent amino acid in a homologue, has been changed to an amino acid with a less polar side chain.

The 'equivalent' side chain in the homologue is one at the homologous position. This can be deduced by lining up the P450_{cam} sequence and the sequence of the homologue based on the homology between the two sequences. The PILEUP, BLAST and BESTFIT algorithms can be used to line up the sequences (for example as described in Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S, F *et al* (1990) J Mol Biol 215:403-10 and (Devereux *et al* (1984) Nucleic Acids Research 12, p387-395)). These algorithms can also be used to calculate the levels of homology discussed herein (for example on their default settings). The equivalent amino acid will generally be in a similar place in the active site of the homologue as amino acid 96 in P450_{cam}.

S4b B² The discussion below provides examples of the positions at which substitutions may be made in P450_{cam}. The same substitutions may be made at equivalent positions in the homologues. Standard nomenclature is used to denote the mutations. The letter of the amino acid present in the natural form is followed by the position, followed by the amino acid in the mutant. To denote multiple mutations in the same protein each mutation is listed separated by hyphens. The mutations discussed below using this nomenclature specify the natural amino acid in P450_{cam}, but it is to be understood that the mutation could be made to a homologue which has a different amino acid at the equivalent position. Note that the amino acid numbering shown in SEQ ID No. 1 for P450_{cam} does not correspond to the numbering used in the description to denote mutations. The numbering in SEQ ID No. 1 is one more than the numbering in the description for a particular position.

-7-

An additional mutation is typically an amino acid substitution at amino acid 87, 98, 101, 185, 244, 247, 248, 296, 395, 396 or a combination of these, for example as shown in table 2.

The following combinations of substitutions are preferred:

5 (i) Substitution at position 87 to amino acids of different side-chain volume, such as substitutions (typically of F) to A, L, I and W, combined with substitutions at position 96 to amino acids of different side-chain volume such as (typically Y to) A, L, F, and W. These combinations alter the space available in the upper part of the substrate pocket compared to the wild-type enzyme, for example, from Y96W-F87W (little space) to
10 Y96A-F87A (more space), as well as the location of the space, for example from one side in Y96F-F87A to the other in Y96A-F87W.

15 (ii) Substitution at position 96 to F combined with substitutions at positions 185 and 395. Both T185 and I395 are at the upper part of the substrate pocket, and substitution with A creates more space while substitution with F will reduce the space available and push the substrate close to the haem.

20 (iii) Substitutions at position 96 to A, L, F, and W combined with substitutions at residues closer to the haem including at 101, 244, 247, 295, 296 and 396 to A, L, F, or W. These combinations will create or reduce space in the region of the different side-chains to offer different binding orientations to substrates of different sizes. For example, the combinations Y96W-L244A and Y96L-V247W will offer very different pockets for the binding of the substrate.

25 (iv) Triple substitutions at combinations of positions 87, 96, 244, 247, 295, 296, 395 and 396 with combinations of A, L, F, and W. The aim is to vary the size and shape of the hydrophobic substrate binding pocket. For example, the Y96A-F87A-L244A combination creates more space compared to the Y96F-F87W-V396L combination, thus allowing larger substrates to bind to the former while restricting the available binding orientations of smaller substrates in the latter. The combinations Y96F-F87W-V247L and Y96F-F87W-V295I have comparable substrate pocket volumes, but the locations of the space available for substrate binding are very different. The combination Y96F-F87L-V247A has a slightly larger side-chain volume at the 96 position than the combination Y96L-F87L-V247A, but the L side-chain at the 96 position is much more flexible and the substrate binding orientations will be different for the two triple mutants.

-8-

(v) The mutants with four or five substitutions were designed with similar principles of manipulating the substrate volume, the different flexibility of various side-chains, and the location of the space available in the substrate pocket for substrate binding so as to effect changes in selectivity of substrate oxidation.

5 Mutations are generally introduced into the enzyme by using methods known in the art, such as site directed mutagenesis of the enzyme, PCR and gene shuffling methods or by the use of multiple mutagenic oligonucleotides in cycles of site-directed mutagenesis. Thus the mutations may be introduced in a directed or random manner. Typically the 10 mutagenesis method produces one or more polynucleotides encoding one or more different mutants. In one embodiment a library of mutant oligonucleotides is produced which can be used to produce a library of mutant enzymes.

The process is typically carried out in the presence of the natural cofactors of the monooxygenase. Thus typically in addition to the enzyme (a) and the substrate the process is carried out in the presence of an electron transfer reductase (b), an electron transfer 15 redoxin (c), cofactor for the enzyme and an oxygen donor. In this system the flow of electrons is generally: cofactor → (b) → (c) → (a).

20 (b) is generally an electron transfer reductase which is able to mediate the transfer of electrons from the cofactor to (c), such as a naturally occurring reductase or a protein which has homology with a naturally occurring reductase, such as at least 70% homology; or a fragment of the reductase or homologue. (b) is typically a reductase of any 25 of the organisms mentioned herein, and is typically a flavin dependent reductase, such as putidaredoxin reductase.

25 (c) is generally an electron transfer redoxin which is able to mediate the transfer of electrons from the cofactor to (a) via (b). (c) is typically a naturally occurring electron transfer redoxin or a protein which has homology with a naturally occurring electron transfer redoxin, such as at least 70% homology; or a fragment of the redoxin or homologue. (c) is typically a redoxin of any of the organisms mentioned herein. (c) is typically a two-iron/two sulphur redoxin, such as putidaredoxin.

30 The cofactor is any compound capable of donating an electron to (b), such as NADH. The oxygen donor is any compound capable of donating oxygen to (a), such as dioxygen.

Typically (a), (b) and (c) are present as separate proteins; however they may be

-9-

present in the same fusion protein. Typically only two of them, preferably (b) and (c), are present in the fusion protein. Typically these components are contiguous in the fusion protein and there is no linker peptide present.

~~546 ③~~ Alternatively a linker may be present between the components. The linker generally comprises amino acids that do not have bulky side chains and therefore do not obstruct the folding of the protein subunits. Preferably the amino acids in the linker are uncharged. Preferred amino acids in the linker are glycine, serine, alanine or threonine. In one embodiment the linker comprises the sequence N-Thr-Asp-Gly-Gly-Ser-Ser-Ser-C. The linker is typically from at least 5 amino acids long, such as at least 10, 30 or 50 or 10 more amino acids long.

In the process the concentration of (a), (b) or (c) is typically from 10^{-8} to $10^{-2}M$, preferably from 10^{-6} to $10^{-4}M$. Typically the ratio of concentrations of (a): (b) and/or (a): (c) is from 0.1:01 to 1:10, preferably from 1:0.5 to 1:2, or from 1:0.8 to 1:1.2. Generally the process is carried out at a temperature and/or pH at which the enzyme is functional, such as when the enzyme has at least 20%, 50%, 80% or more of peak activity. Typically the pH is from 3 to 11, such as 5 to 9 or 6 to 8, preferably 7 to 7.8 or 7.4. Typically the temperature is 10 to 90°C, such as 25 to 75°C or 30 to 60°C.

In the process different monooxygenases may be present. Typically each of these will be able to oxidise different substrates, and thus using a mixture of monooxygenases will enable a wider range of substrates to be oxidised.

In one embodiment the process is carried out in the presence of a substance able to remove hydrogen peroxide by-product (e.g. a catalase).

In one embodiment the process is carried out in the presence of the enzyme, substrate and an oxygen atom donor, such as hydrogen peroxide or t-butylhydroperoxide, for example using the peroxide shunt.

In one embodiment in the process the (a), (b) and (c) together are typically in a substantially isolated form and/or a substantially purified form, in which case together they will generally comprise at least 90%, e.g., at least 95%, 98% or 99% of the protein in the preparation.

The process may be carried out inside or outside a cell. The cell is typically in culture, at a locus, in vivo or in planta (these aspects are discussed below).

The process is typically carried out at a locus such as in land (e.g. in soil) or in

-10-

water (e.g. fresh water or sea water). When it carried out in culture the culture typically comprises different types of cells of the invention, for example expressing different monooxygenases of the invention. Generally such cells are cultured in the presence of assimible carbon and nitrogen sources.

5 Typically the cell in which the process is carried out is one in which the monooxygenase does not naturally occur. In another embodiment the monooxygenase is expressed in a cell in which it does naturally occur, but at higher levels than naturally occurring levels. The cell may produce 1, 2, 3, 4 or more different monooxygenases of the invention. These monooxygenases may be capable of oxidising different halo aromatic
10 compounds. Typically the cell also expresses any of the reductases and/or redoxins discussed above.

The cell is typically produced by introducing into a cell (i.e. transforming the cell with) a vector comprising a polynucleotide that encodes the monooxygenase. The vector may integrate into the genome of the cell or remain extrachromosomal. The cell may
15 develop into the animal or plant discussed below. Typically the coding sequence of the polynucleotide is operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell. The control sequence is generally a promoter, typically of the cell in which the monooxygenase expressed.

20 The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

25 The vector is typically a transposon, plasmid, virus or phage vector. It typically comprises an origin of replication. It typically comprises one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid. The vector is typically introduced into host cells using conventional techniques including calcium phosphate precipitation, DEAE-dextran transfection, or electroporation.

30 Components (b) and (c) may be expressed in the cell in a similar manner. Typically (a), (b) and (c) are expressed from the same vector, or may be expressed from different vectors. They may be expressed as three different polypeptides. Alternatively they may be expressed in the form of fusion proteins. The cell typically expresses more

-11-

than one type of monooxygenase.

Sub C4 In one embodiment the three genes encoding the three proteins of the P450_{cam} system, i.e. camA, camB, and camC are placed in the mobile regions of standard transposon vectors and incorporated into the genome of Pseudomonas and flavobacteria.

5 Alternatively plasmid vectors for expressing these genes may be used, in which case the P450_{cam} gene cluster will be extra-chromosomal.

The cell may be prokaryotic or eukaryotic and is generally any of the cells or of any of the organisms mentioned herein. Preferred cells are Pseudomonas, flavobacteria or fungi cells (e.g. Aspergillus). In one embodiment the cell is one which in its naturally occurring form is able to oxidise any of the substrates mentioned herein. Typically the cell is in a substantially isolated form and/or substantially purified form, in which case it will generally comprise at least 90%, e.g. at least 95%, 98% or 99% of the cells or dry mass of the preparation.

10 The invention provides a transgenic animal or plant whose cells are any of the cells of the invention. The animal or plant is transgenic for the monooxygenase gene and typically also an appropriate electron transfer reductase and/or redoxin gene. They may be homozygote or heterozygote for such genes, which are typically transiently introduced into the cells, or stably integrated (e.g. in the genome). The animal is typically a worm (e.g. earthworm) or nematode. The plant or animal may be obtained by transforming an appropriate cell (e.g. embryo stem cell, callus or germ cell), fertilising the cell if required, allowing the cell to develop into the animal or plant and breeding the animal or plant true if required. The animal or plant may be obtained by sexual or asexual (e.g. cloning) propagation of an animal or plant of the invention or of the F1 organism (or any generation removed from the F1, or the chimera that develops from the transformed cell).

15 25 As discussed above the process may be carried out at a locus. Thus the invention also provides a method of treating a locus contaminated with a halo aromatic compound comprising contacting the locus with a monooxygenase, cell, animal or plant of the invention. These organisms are then typically allowed to oxidise the halo aromatic compound. In one embodiment the organisms used to treat the locus are native to the locus. Thus they may be obtained from the locus (e.g. after contamination), transformed/transfected (as discussed above) to express the monooxygenase (and optionally an appropriate electron transfer reductase and/or redoxin).

-12-

In one embodiment the locus is treated with more than one type of organism of the invention, e.g. with 2, 3, 4, or more types which express different monooxygenases which oxidise different halo aromatic compounds. In one embodiment such a collection of organisms between them is able to oxidise all halobenzenes, e.g. all chlorobenzenes.

5 The organisms (e.g. in the form of the collection) may carry out the process of the invention in a bioreactor (e.g. in which they are present in immobilised form). Thus the water or soil to be treated may be passed through such a bioreactor. Soil may be washed with water augmented with surfactants or ethanol and then introduced into the bioreactor.

10 The invention also provides a process for selecting a mutant of a monooxygenase for its ability to oxidise any of the substrates mentioned herein, which process comprises screening a library of said mutants for their oxidation effect on the substrate. Thus typically the substrate is provided to the library and mutants are selected based on their ability to oxidise the substrate, for example at a particular rate or under particular conditions. The mutant may be selected based on its ability to oxidise the substrate to a 15 particular oxidation product.

Typically the library will be in the form of cells which comprise the mutant enzymes. Generally each cell will express only one particular mutant enzyme. The library typically comprises at least 500 mutants, such as at least 1,000 or 5,000 mutants, preferably at least 10,000 different mutants.

20 The library typically comprises a random population of mutants. The library may undergo one or more rounds of selection whilst being produced and therefore may not comprise a random population.

25 The library is typically produced by contacting any of the cells discussed herein which expresses the monooxygenase with a mutagen and/or when the cell is a mutator cell culturing the cell in conditions in which mutants are produced. The mutagen may be contacted with the cell prior to or during culturing of the cell. Thus the mutagen may be present during replication of the cell or replication of the genome of the cell.

30 The mutagen generally causes random mutations in the polynucleotide sequence which encodes (a). The mutagen is typically a chemical mutagen, such as nitrosomethylguanidine, methyl- or ethylmethane sulphonic acid, nitrite, hydroxylamine, DNA base analogues, and acridine dyes, such as proflavin. It is typically electromagnetic radiation, such as ultra-violet radiation at 260 nm (absorption maximum of DNA) and X-

-13-

rays. It is typically ionising radiation.

A mutator cell is generally deficient in one or more of the primary DNA repair pathways (such as *E.Coli* pathways mutS, mutD or mutT, or their equivalents in another organism), and thus has a high mutation rate. Simply culturing such cell leads to the DNA encoding (a) to become mutated. The cell may be of *E.Coli* XL1 Red mutator strain.

The mutant selected from the library may be used in any aspect of the invention, thus it may be used to oxidise a substrate in the process of the invention or may be expressed in the cell, animal or plant of the invention. It may be used in the method of treating a locus.

10 The invention is also illustrated by the Examples:

Example 1

Expression of mutants for *in vitro* work.

The P450_{cam} enzymes were expressed using the vector pRH1091 (Baldwin, J.E., Blackburn, J.M., Heath, R.J., and Sutherland, J.D. *Bioorg. Med. Chem. Letts.*, 1992, 2, 663-668.) which utilised the *trc* promoter (a fusion of the *trp* and *lac* promoters). This vector incorporates a strong ribosome binding site (RBS) and the gene to be expressed is cloned using an *Nde* I site on the 5' end of the gene. We used *Hind* III as the cloning site at the 3' end of the *camC* gene. The procedure for protein expression is as follows: Cells are grown at 30°C until the OD_{600nm} reaches 1.0 – 1.2, the temperature is increased to 37°C and camphor added as a 1 M stock in ethanol to a final concentration of 1 mM. The culture is allowed to incubate at 37°C for another 6 hours. The P450_{cam} protein is expressed to high levels in the cytoplasm and the cells take on a red to orange-red colour.

We have also prepared a variant of pRH1091 (by PCR) which has an extra *Xba* I site between the RBS and the *Nde* I site. This is important because *Nde* I is not unique in M13, and this restriction site is also present in the reductase gene as well as the backbone of the pGLW11 vector used for the *in vivo* system. *Xba* I is unique in the polylinker region of M13, but absent in the genes of all three proteins in the P450_{cam} system and in the expression vectors. It therefore allows the *camC* gene to be moved between the mutagenic and expression vectors.

-14-

Svb 85 H w the mutants were made.

Oligonucleotide-directed site-specific mutagenesis was carried out by the Kunkel method (Kunkel, T. A. *Proc. Natl. Acad. Sci. USA* 1985, 82, 488-492) using the Bio-Rad Mutagen kit. The recommended procedure is summarised as follows. An M13 mp19 subclone of the *camC* gene encoding P450_{cam} was propagated in the *E. coli* strain CJ236. This strain has the *dut*⁻ *ung*⁻ phenotype and thus will tolerate the inclusion of uracil in place of thymine in DNA molecules. After three cycles of infection, uracil-containing single stranded (USS) M13 DNA was readily isolated by phenol extraction of mature M13 phage particles excreted into the growth medium. The mutagenic oligonucleotide (or oligonucleotides) were phosphorylated with T4 polynucleotide kinase and then annealed to the USS template. The four nucleotides, DNA polymerase, DNA ligase, ATP and other chemical components were added and the second strand was synthesised *in vitro*. The double stranded form thus obtained was transformed into the *dut*⁺ *ung*⁺ *E. coli* strain MV1190, which should degrade the uracil-containing template strand and propagate the mutant strand synthesised *in vitro*. Plaques were picked and phages of possible mutants grown in *E. coli* strains MV1190 or TG1. The single-stranded DNA from these were sequenced to determine whether the mutagenesis reaction was successful. The mutagenic efficiency was 50 - 80%.

The mutant *camC* gene is excised from the M13 subclone by restriction digest with *Nde* I and *Hind* III, and the fragment of appropriate size is ligated to the backbone of the expression vector prepared by a similar *Nde* I/*Hind* III digest.

Multiple mutants were prepared either by further mutagenesis, also by the Kunkel method, or where the location of the sites in the sequence permits, simple cloning steps. There are two unique restriction sites within the *camC* gene which are absent from the expression vector. One is *Sph* I which spans residues 121 - 123, and the other is *Sal* I which spans residues 338 and 339. Therefore, all mutations at, for example, residues 87, 96, 98, and 101 are readily combined with mutations at higher number residues by ligating appropriate fragments from restriction digests of mutant *camC* genes with *Nde* I/*Sph* I and *Sph* I/*Hind* III and the backbone fragment from a *Nde* I/*Sph* I digest of the expression vector. Mutations at, for example, 395 and 396 can be similarly incorporated by digests in which *Sph* I is replaced with *Sal* I.

The rationale for introducing the unique *Xba* I site is now clear: many mutants with

-15-

multiple mutations were prepared by the cloning procedure above. Without the *Xba* I site it would be impossible to clone the gene for these multiple mutants from the expression vector back into M13 for further rounds of mutagenesis. Of course these problems could be overcome by doing mutagenesis by PCR, for example.

5 Example 2Substrate oxidation protocol: *in vitro* reactions

Component	Final concentration
P450 _{cam} enzyme	1 µM
Putidaredoxin	10 µM
Putidaredoxin reductase	1 µM
Bovine liver catalase	20 µg/ml
KCl	200 mM
Substrate	Typically 1 mM
NADH	250 - 400 µM

15 * 50 mM Tris-HCl buffer pH 7.4 is added to make up the volume.
 * Temperature controlled at 30°C, optional.
 * The NADH turnover rate could be determined by monitoring the absorbance at 340 nm with time.
 * Catalase does not catalyse the substrate oxidation reactions but rather it is present

20 to remove any hydrogen peroxide by-product which could otherwise denature the P450_{cam}.

The method can be increased in scale to, for example, 20 ml total incubation volume to allow purification of sufficient products by HPLC for spectroscopic characterisation. Fresh substrate (1 mM) and NADH (1 - 2 mM) are added periodically, such as every 20 minutes in a total reaction time of, typically, 3 hours.

25 Example 3The *in vivo* system

The *in vivo* systems were expressed using the vector pGLW11, a derivative of the plasmid pKK223 (Brosius, J. and Holy, A. *Proc. Natl. Acad. Sci. USA*, 1984, 81, 6929-6933). Expression is directed by the *tac* promoter and the vector incorporates a gene

-16-

conferring resistance to the antibiotic ampicillin.

Two systems were constructed. The first one expressed the electron transfer proteins putidaredoxin reductase (*camA gene*) and *putidaredoxin* (*camB gene*) as a fusion protein with a seven amino acid peptide linker, and the P450_{cam} enzyme (*camC gene*) was expressed by the same vector but it was not fused to the electron proteins. The second system expressed the three proteins as separate entities in the *E.Coli* host. Both systems were catalytically competent for substrate oxidation *in vivo*.

The general strategy was as follows. The genes for the three proteins were cloned using *Eco RI* and *Hind III* as flanking sites, with *Eco RI* at the 5' end. For both *in vivo* systems there are restriction sites between the genes, including between the reductase and redoxin genes in the fusion construct. These restriction sites were introduced by PCR, as detailed below. The first task, however, was to carry out a silent mutation to remove the *Hind III* site within the *camA* gene for the reductase. The AAGCTT *Hind III* recognition sequence in the *camA* gene was changed to AAGCCT, which is a silent mutation because GCT and GCC both encode alanine. The gene was completely sequenced to ensure that there were no spurious mutations.

1. The fusion protein system

1.a Manipulation of the *camA* gene by PCR

Sub B For the *camA* gene the primer below was used at the 5' end of the gene to introduce the *Eco RI* cloning site and to change the first codon from GTG to the strong start codon ATG.

5'-GAG ATT AAG AAT TCA TAA ACA CAT GGG AGT GCG TGC CAT ATG AAC GCA AAC
Eco RI RBS |-*camA*

Sub B 7 At the 3' end of *camA* the primer was designed such that 15 bases are complementary to nucleotide sequence of the last five amino acid residues of *camA*. The stop codon immediately after the GCC codon for the last amino acid was removed, and then part of a seven amino acid linker (Thr Asp Gly Gly Ser Ser Ser) which contained a *Bam HI* cloning site (GGATCC = Gly Ser) was introduced. The coding sequence was thus:

-17-

5'- GAA CTG AGT AGT GCC ACT GAC GGA **GGA TCC TCA TCG-3'**

camA → Thr Asp Gly Gly Ser

|*Bam* HI|

Sub B⁸

5' The primer sequence shown below is the reverse complement used for PCR:

5'- CGA TGA **GGA TCC TCC GTC AGT GGC ACT ACT CAG TTC-3'**

Sub B⁹

1.b Manipulations of the *camB* gene by PCR

For the *camB* gene the primer at the 5' end incorporated the second half of the peptide linker between the reductase and redoxin proteins, and the restriction site *Bam* HI for joining the two amplified genes together.

5'- TCA TCG GGA TCC TCA TCG ATG TCT AAA GTA GTG TAT-3'

Gly Ser Ser Ser | → *camB*

|*Bam* HI| Start

Sub B¹⁰

At the 3' end of *camB* the primer incorporates 12 nucleotides complementary to the end of *camB* followed by the stop codon TAA, a 6 nucleotide spacer before the GGAG ribosome binding site. *Xba* I and *Hind* III sites were then added to allow cloning of the *camC* gene when required. The sequence of the coding strand was therefore:

5'- CCC GAT AGG CAA TGG **TAA TCA TCG GGAG TCT AGA GCA TCG AAG CTT TCA TCG-3'**

CamB → |stop RBS *Xba* I *Hind* III

Sub B¹¹

20 The primer shown below is the reverse complement used for PCR:

5'-CGA TGA AAG CTT CGA TGC TCT AGA CTCC CGA TGA TTA CCA TTG CCT ATC GGG -3'

1.c Preparation of the full fusion construct

The *camA* and *camB* genes were amplified by the PCR using the primers described above. The new *camA* was digested with *Eco* RI and *Bam* HI, while the new *CamB* was digested with *Bam* HI and *Hind* III. The pGLW11 expression vector was digested with *Eco* RI and *Hind* III. All three were purified by agarose gel electrophoresis and the three gel slices containing the separate fragments were excised from the gel and ligated together, and then transformed into *E.Coli* DH5α. Successful ligation of all the fragments were

-18-

confirmed by a series of restriction digestion experiments, especially the presence of the new and unique *Xba* I site. The entire sequence of the insert from the *Eco* RI site to the *Hind* III site was determined to ensure that all the sequences were correct.

The new plasmid, named pSGB^F, was transformed into *E.Coli* and expression of the reductase and redoxin proteins was induced by IPTG. When a purified P450_{cam} enzyme was added to the cell-free extract, substrate oxidation was observed for a variety of substrates.

When the *camC* gene is cloned into the pSGB^F plasmid using the *Xba* I and *Hind* III restriction sites, the new recombinant plasmid thus generated expresses the reductase and redoxin as a fusion protein and the P450_{cam} enzyme as a operate entity both from the same mRNA molecule. This *in vivo* system is catalytically competent for terpene oxidation in whole cells.

2. The *in vivo* system with the protein expressed separately

2.a The basic strategy

The starting point of the preparation of this *in vivo* system was the recombinant plasmid used to express the *camA* gene for putidaredoxin reductase. The *camA* gene was cloned into the pGLW11 plasmid using the *Eco* RI and *Bam* HI restriction sites, with *Eco* RI being at the 5' end of the gene. Conveniently the polylinker region of the pGLW11 vector has a *Hind* III site downstream of the *Bam* HI site. The *camB* gene was therefore manipulated by PCR such that it can be cloned into pGLW11 using the *Bam* HI and *Hind* III sites. This new plasmid expresses the reductase and redoxin as separate proteins.

The *camB* gene was cloned into pUC118 by the *Bam* HI and *Hind* III cloning sites to express putidaredoxin for our general *in vitro* substrate oxidation work. Therefore, the PCR primer at the 3' end of the *camB* gene was designed to introduce a ribosome binding site and the *Xba* I restriction site upstream of the *Hind* III site so that the *camC* gene can be inserted downstream of *camB* using the *Xba* I and *Hind* III sites. Therefore the three genes were cloned without fusion in the pGLW11 expression vector and arranged in the order 5'-*camA-camB-camC-3'*, and each gene has its own RBS to initiate protein synthesis.

-19-

2.b Manipulations of the *camB* gene

54b B¹² We used the internal and unique restriction site *Mlu* I (recognition sequence ACGCGT) within the *camB* gene as the starting point so that the PCR product has a different size from the PCR template fragment. The primers were as follows:

5 5'- TCA TCG ACG **CGT** CGC GAA CTG CTG-3'

where the *Mlu* I site is in bold.

54b B¹³ The desired coding sequence at the 3' end of the *camB* gene was:

5'- CCC GAT AGG CAA TGG TAA GTA GGT GAA TAT **CTA** ATC CCC ATC
camB - | stop

10 10 TAT GCG CGA GTG **GAG** TCT AGA GTT CGA-3'

RBS Xba I

After the stop codon there is a 35 base spacer before the RBS which is used to initiate the synthesis of the P450_{cam} enzyme. The *Xba* I cloning site is located within the spacer between the RBS and the start codon (not in this primer) of the *camC* gene. The 15 PCR primer used was the reverse complement of the sequence above. The PCR was carried out and the amplified fragment of the appropriate size was purified by agarose gel electrophoresis and the gel slice excised.

One extra step was necessary to complete the construction of the new plasmid. The 20 plasmid for the fusion protein *in vivo* system was digested with *Mlu* I and *Hind* III restriction enzymes, purified by agarose gel electrophoresis, and the gel slice for the small 25 *camB* fragment excised. The pUC118 plasmid for *camB* expression was similarly digested, and the gel slice for the backbone was excised. By ligating the two fragments together we prepared a new pUC118-based plasmid which had an *Xba* I site followed by an *Hind* III site downstream of the stop codon of *camB*. This new plasmid was digested with the *Mlu* I and *Xba* I enzymes and the backbone was ligated with the new *camB* fragment described above to generate a plasmid with the following arrangement of the key components:

-20-

..*lac* Promoter..*Bam* HI..*camB* gene..spacer..RBS..*Xba* I..*Hind* III..

2.c Preparation of the *in vivo* system plasmid

Once the modified *camB* with the *Xba* I and *Hind* III restriction sites and appropriate spacers were prepared, the *in vivo* system was constructed by cloning this into the pGLW11-based plasmid used to express the *camA* gene (reductase protein) using the *Bam* HI and *Hind* III sites. The new *in vivo* system vector has the following arrangement of the key components:

..*tac* Promoter..*Eco* IRI..RBS..*camA* gene..spacer..*Bam* HI..RBS..*camB* gene..spacer..RBS..*Xba* I..*Hind* III..

This new plasmid, named pSGB⁺, was transformed into *E.Coli* and expression of the reductase and redoxin proteins was induced by IPTG. When a purified P450_{cam} enzyme was added to the cell-free extract, substrate oxidation was observed for a variety of substrates.

When the *camC* gene is cloned into this pSGC⁺ plasmid using the *Xba* I and *Hind* III restriction sites, the new recombinant plasmid thus generated will express the three proteins separately, each under the direction of its own RBS but from the same mRNA molecule. Thus constitutes the *in vivo* system used in the vast majority of our terpene oxidation work.

3. Introduction of an *Xba* I site into pRH1091

This is the final step to enable the *camC* gene to be cloned into the *in vivo* systems by the two cloning sites *Xba*I and *Hind* III. The *Xba* I site was added by PCR of the entire pRH1091 plasmid using two primers. The presence of these two sites will also enable cloning of the *camC* gene into M13 since both *Xba* I and *Hind* III are unique in *camC* and M13.

Sub 25 *B* 14
The primers shown below maintain the *Hind* III cloning site AAGCTT:

5'-TCA TCG AAG CTT GGC TGT TTT-3'

Hind III | - vector

-21-

At the other end the coding sequence desired was:

5'-ACA ATT TCA CAC AGGA TCT AGA C CAT ATG TCA TCG AAG CTT TCA TCG-3'

Vector - |RBS *Xba* I *Nde* I *Hind* III

This sequence maintained the *Nde* I and *Hind* III sites but the new *Xba* I site was
 5 introduced upstream of the *Nde* I site. The PCR primer used was the reverse complement
 of the desired sequence:

5'-CGA TGA AAG CTT CGA TGA CAT ATG GTC T AGA TCCT GTG TGA AAT TGT-3'

The PCR product was then purified by agarose gel electrophoresis, digested with
Hind III and circularised with T4 DNA ligase. Success of the PCR method was indicated
 10 by the presence of a new and unique *Xba* I site in plasmid DNA isolated from
 transformants.

4. Cloning of *camC* into the *in vivo* systems

All existing *camC* mutants were cut out of pRH1091-based expression plastids with
Nde I and *Hind* III. The new vector is similarly cut with the same restriction enzymes and
 15 the *camC* gene cloned into this plasmid with T4 DNA ligase. This DNA is transformed
 into *E.Coli* JM109 which then may be grown to express P450_{cam}.

The *camC* gene is excised from the new vector using *Xba* I and *Hind* III restriction
 enzymes and cloned into either the *in vivo* vector systems or M13mp19 for mutagenesis.

20 5. *In vivo* expression and substrate turnover

For protein expression, cells are grown in LBamp medium (tryptone 10 g/litre,
 yeast extract 5 g/litre, NaCl 10 g/litre, 50 µg/ml ampicillin) at 30°C until the OD_{600nm}
 reaches 1.0 - 1.2. IPTG (isopropyl-β-D-thiogalactopyranoside) was added to a final
 concentration of 1 µM (from a 1 M stock in H₂O) and the culture was incubated at 30°C
 25 overnight.

For simple screening the substrate can be added to culture and the incubation
 continued. However, due to impurities from the culture media the cells were generally
 washed twice with 0.5 vol. of buffer P, (KH₂PO₄ 6.4 g, K₂HPO₄.3H₂O 25.8 g, H₂O to 4

-22-

litres, pH 7.4) and resuspended in 0.25 vol. oxygen saturated buffer P containing 24 mM glucose. Substrate was added to 1 mM and the incubation continued at 30°C. The reaction was allowed to run for 24 hours with periodic additions of substrate and glucose.

Example 4

The oxidation of halo aromatic compounds

Mutant	2,3,6-Trichlorophenol	3,4,6-Trichlorophenol	Coupling Efficiency (%)	Product formation rate (min ⁻¹)
Y96F	75	25	18	22
Y96A	77	23	14	33
Y96H	54	46	3	1
F87L-Y96F	42	58	4	8
F87A-Y96F	52	48	2	3
F87A-Y96-F-V247A	43	57	4	7

Mutant	2,3-Dichlorophenol	3,4-Dichlorophenol	Coupling Efficiency (%)	Product formation rate (min ⁻¹)
Y96A	94	6	6	19
Y96F	91	9	4	8
Y96A-V247L	94	6	7	20
Y96L-V247A	90	10	2	0.7
F87L-96F	96	4	3	5
C334A	95	5	2	0.5

All mutants have C334A. Coupling efficiency is the percentage of NADH consumed which was utilised for product formation, i.e. a percentage of the theoretical maximum efficiency. The product formation rates are given in (nmol product) (nmol P450_{cum})⁻¹ (min)⁻¹. The relative amount of product formed in each case is shown.

-23-

1,3- and 1,4-dichlorobenzene, 1, 2, 3- and 1, 3, 5-trichlorobenzene, 1, 2, 4, 5- and 1, 2, 3, 5-tetrachlorobenzene, and 2, 3, 4, 5, 6- and 2, 2', 4, 5, 5'- pentachlorobiphenyl were also found to be oxidised.

Wild-type and mutant P450_{cam} enzymes were tested for their ability to oxidise 3,3'-dichlorobiphenyl and 2,2',4,5,5'-pentachlorobiphenyl. Results are shown in terms of NADH turnover. Rates are given as nanomol NADH consumed per nanomol P450_{cam} enzyme per minute.

	P450_{cam} enzyme	3,3'-dichlorobiphenyl	2,2',4,5,5'-pentachlorobiphenyl
10	Wild-type	0.4	not detected
	Y96F	15	1
	F87A-Y96F	845	165
	F87L-Y96F	174	13
	F87W-Y96F	4	3
15	F87A-Y96F-V247A	112	12
	Y96A-V247L	84	37
	F87A-Y96F-L244A	669	321
	F87A-Y97F-L244A-V247A	173	214

The first product, 4-hydroxy-3,3'-dichlorobiphenyl was identified by the characteristic coupling patterns expected in the ¹H NMR spectrum and by mass spectroscopy. The further oxidation product, 4,4'-dihydroxy-3,3'-dichlorobiphenyl was identified by co-elution with an authentic sample, and by UV-vis and mass spectroscopy. This product did not constitute more than ca. 10% of the total products in any of the mutants tested.

For the second substrate product was established as 4'-hydroxy-2,2',4,5,5'-pentachlorobiphenyl by the observation of the parent ion in the mass spectrum, and by comparison with literature ¹H NMR data.

-24-

P450_{cam} mutants

All mutants optionally contain the base mutation C334A.

Single mutants: Y96A, Y96F, Y96L, Y96W.

Double mutants:

Y96A-F87A	Y96F-F87A	Y96F-V295A	Y96L-F87A	Y96L-A296L
Y96A-F87L	Y96F-F87I	Y96F-V295L	Y96L-F87L	Y96L-A296F
Y96A-F87W	Y96F-F87L	Y96F-V295I	Y96L-F98W	Y96L-V396A
Y96A-F98W	Y96F-F87W	Y96F-A296L	Y96L-T101L	Y96L-V396L
Y96A-L244A	Y96F-F98W	Y96F-A296F	Y96L-T101F	Y96L-V396F
Y96A-V247A	Y96F-T101L	Y96F-I395F	Y96L-L244A	Y96L-V396W
Y96A-V247L	Y96F-T101F	Y96F-I395G	Y96L-L244F	
Y96A-I395F	Y96F-T185A	Y96F-V396A	Y96L-V247A	
Y96A-I395G	Y96F-T185F	Y96F-V396L	Y96L-V247L	Y96W-F87W
	Y96F-T185L	Y96F-V396F	Y96L-V247F	Y96W-F98W
	Y96F-L244A	Y96F-V396W	Y96L-V247W	Y96W-L244A
	Y96F-V247A		Y96L-G248L	Y96W-V247A
	Y96F-V247L		Y96L-V295L	Y96W-V396A
	Y96F-G248L		Y96L-V295F	

Triple Mutants:

Y96A-F87A-L244A	Y96L-V247A-V396L	Y96F-F87W-V247A
Y96A-F87A-V247A	Y96L-V247A-V396F	Y96F-F87W-V247L
Y96A-F87L-L244A	Y96L-V247A-V396W	Y96F-F87W-V247F
Y96A-F87L-V247A	Y96L-V247F-V396A	Y96F-F87W-V295L
Y96A-L244A-V247A		Y96F-F87W-A296L
	Y96F-F87A-L244A	Y96F-F87W-V396A
Y96L-F87A-L244A	Y96F-F87A-V247A	Y96F-F87W-V396L
Y96L-F87A-V247A	Z96F-F87A-V247L	Y96F-V247F-V396A
Y96L-F87L-L244A	Y96F-F87A-I395F	Y96F-L244A-V396L
Y96L-F87L-V247A	Y96F-F87A-I395G	Y96F-L244A-V396F
Y96L-V247A-I395F	Y96F-F87L-V247A	Y96F-L244A-V396W
Y96L-V247L-I395F	Y96F-F87L-V247L	Y96F-L244F-V396A
Y96L-V247L-I395G	Y96F-F87L-I395F	Y96F-V247A-V396L
Y96L-L244A-V396L	Y96F-F87W-T185A	Y96F-V247A-V396F
Y96L-L244A-V396F	Y96F-F87W-T185F	Y96F-V247A-V396W
Y96L-L244A-V396W	Y96F-F87W-T185L	
Y96L-L244F-V396A	Y96F-F87W-L244F	Y96W-F87W-F98W

Four mutations:

Y96A-F87A-L244A-V247A
 Y96A-F87L-L244A-V247A
 Y96L-F87A-L244A-V247A
 Y96L-F87L-L244A-V247A

Five mutations:

Y96F-F87W-T185L-V247L-V295L
 Y96F-F87W-T185L-V247L-V396A
 Y96F-F87W-T185L-V247L-V396L
 Y96L-F87L-L244A-V247A

Y96F-F87W-L244A-V295L
 Y96F-F87W-L244F-V396A
 Y96F-F87W-L244A-A296L
 Y96F-F87W-V247A-V396L
 Y96F-F87W-V247A-V396F
 Y96F-F87W-V247L-V295A
 Y96F-F87W-V247L-V396A
 Y96F-F87W-V247F-V396A
 Y96F-F87W-V247A-I395F
 Y96F-F87W-V247L-I395G